5 Role of genes in CFA

The genetics of craniofacial anomalies and of cleft lip and palate, in particular, as the single most important sentinel defect of this group is highly complex. As is evident from this summary report and others that accompany it, etiologies are many-fold and complex and include single-gene causes, chromosomal disorders, polygenic interactions, environmental risks, gene/environment risks, and even the likely role of chance. Studies in this area began formally in the 1930s and 1940s with the work of Paul Fogh-Andersen and subsequently continued in the 1950s with Clark Fraser (1968). In the ensuing years much has been learned about the genetics of craniofacial anomalies and the recent advances in the progress of the Human Genome Project with the availability of almost complete human and mouse sequence provide unique and special opportunities to further these studies in powerful ways.

At the same time that the genetics is advancing, it is also clear that many questions remain, including even basic questions of phenotype definition and strategies for gene identification. Equally importantly, these studies need to be carried out in conjunction with other investigators whose primary interests and abilities lie in the areas of epidemiology, environment, nutrition, and clinical trials and prevention. The success of folic acid interventions in preventing neural tube defects provides a benchmark against which other preventive strategies for birth defects can be measured and the hope is that improvements in surgical techniques, speech pathology, dental care, nursing, psychological and paediatric care, and the many other fields involved with children with CFA will occur in concert with studies of etiology and prevention. By working together we can all provide a better future for children born with CFA, in the hope that prevention of these defects occurring in children will also be soon on the horizon.

5.1 Embryogenesis

By working together we can all provide a better future for children born with CFA

Development of craniofacial structures represents the complex interactions of many genes and environmental triggers. Studies of monozygotic twins whose facial appearances are almost completely overlapping in recognizable phenotypic features tell us that the role of genetics is almost 100% determinant in providing the outline of normal facial structures. Similarly, studies of monozygotic twins show a much higher concordance rate for non-syndromic forms of cleft lip and palate than would be found in dizygotic twins or siblings, again supporting the strong role of genetics in the etiology of defects of development. Nonetheless, concordance is only between 40% and 60% for clefting in monozygotic twins, which strongly supports the observation that the role of *in utero* environment or possibly some element of stochastic variation is also critical in determining which child might be born with which particular form of craniofacial disorder. Independent of non-syndromic forms of cleft lip and palate are many other defects of craniofacial regions, including other forms of clefts, craniosynostosis, and ocular and ear anomalies that have equally wide and disparate causes. During the course of the meeting, the etiology and pathogenesis of both orofacial clefting and craniosynostosis were reviewed in detail by Dr Michael Cohen (Cohen, 1995). In addition, the entire topic of craniofacial development has recently been extensively reviewed and reported upon by Geoffrey Sperber (2001), and this text as well as other recent publications on embryogenesis can serve as valuable resources for individuals with an interest in craniofacial disorders. Recent references include extensive lists of genes that have already been shown to play an important role in facial development; these genes can, in many cases, be divided into the roles that they play in a variety of morphogenetic pathways. These can include genes identified as growth factors, cytokines, self-signalling molecules, structural proteins (such as collagens or extra cellular matrix proteins), and other forms of morphogens or signalling molecules. Table 6 below lists a few of the genes that, from available genetic evidence, play a role in facial development; this list is in no way comprehensive and is, in fact, changing almost daily.

Table 6: Identified genes/clefts

Genes	Syndromes
CDKN1C	Beckwith-Wiedemann
COL11A1	Marshall
COL11A2	Stickler/Nance-Insley
COL2A1	Stickler/Kniest
CREBBP	Rubinstein-Taybi
DHCR7	Smith-Lemli-Opitz
DTDST	Diastrophic dysplasia
FGD1	Aarskog
FGFR2	Apert
FKHL15	Hypothyroidism
GLI3	Grieg/Pallister-Hall
GPC3	Simpson-Golabi-Behmel
KAL1	Kallman
L1CAM	MASA
LMX1B	Nail-patella
MID1	Opitz
MITF	Waardenburg 2A
PAX3	Waardenburg
PEX1,2,5,6,12	Zellweger
PTCH	Basal cell nevus
SHH	Holoprosencephaly
SIX3	Holoprosencephaly
SOX9	Campomelic dysplasia
TREACLE	Treacher Collins
TWIST	Saethre-Chotzen

Source: Murray JC, 2002

The availability of web sites provides opportunities to update the ongoing lists of candidates, as do the databases of clinical disorders involving craniofacial structure; these databases now identify many hundreds of such disorders. Valuable web sites for discussions of clinical aspects of human craniofacial disorders are listed below.

Table 7: Web sites

OMIM http://www3.ncbi.nlm.nih.gov/Omim/searchomim.html
A listing of Mendelian disorders and genes; comprehensive for humans and extensively referenced with descriptive and historical data.
Human dysmorphology database http://www.hgmp.mrc.ac.uk/DHMHD/dysmorph.html
A searchable database that provides both human and mouse homologies and also allows identification of disorders based on clinical, phenotypic and laboratory features.
Gene Tests/ Gene Clinics http://www.geneclinics.org/
Gene Tests and Gene Clinics are complementary databases. Gene Clinics provides descriptions of many genetic disorders, with an emphasis on management and diagnosis. Gene Tests provides a listing of both clinical and research laboratories currently carrying out molecular studies on a wide range of human disorders, including those involving craniofacial structures.

5.2 Clinical definition of craniofacial anomalies

This topic included discussions by Drs Michael Cohen, Marilyn Jones and Howard Saal of how craniofacial disorders can be defined from a broad perspective, with focused discussions on what would constitute the difference between non-syndromic and syndromic forms of cleft lip and palate and cleft palate only (Jones, 1988).

From the perspective of syndromic identification, many syndromes are now undergoing a revolution in their description as molecular ab-normalities of individual genes are defined and redefined. This has been particularly evident in the description of the craniosynostosis syndromes as a variety of fibroblast growth-factor receptor genes, as well as at least one homeobox gene, have been demonstrated as having mutations that are etiologic for those disorders previously described as phenotypes. The situation has become immediately complex with different genes demonstrating mutations with apparently similar phenotypes, such as Pfeiffer syndrome associations with both FGFR1 and FGFR2 mutations, as well as the same gene having mutations that would have been separated on the basis of phenotypic appearances, such as Crouzon's and Pfeiffer's and mutations in FGFR2. Extensive discussions regarding the role that molecular definitions should play in conjunction with clinical delineation took place.

From the perspective of non-syndromic forms of clefting, the discussion was equally wide-ranging. Historically, based on animal as well as human segregation analysis and recurrence risk studies, cleft lip with or without cleft palate has been separated from cleft palate only. It is now evident

that there can be at least occasional overlap between these phenotypes, as has been demonstrated for MSX1 mutations in at least one large family that includes individuals with isolated cleft palate, as well as cleft lip and palate (van den Boogaard, 2000). It has also been recognized for the last few decades in the case of the autosomal-dominant van der Woude's syndrome. Thus, the historic separation of these two categories on embryologic and genetic grounds – while still a valuable tool – is not 100% representative of observational data.

The description of what constitutes non-syndromic forms of clefting was also extensive and has yet to be fully resolved. This discussion is important in that studies undertaking genetic mapping of cleft lip and palate have increased power when phenotypes can be accurately and reproducibly identified. Thus, the ability to generate sub-phenotypes based upon what might have previously been thought of as normal variation is especially critical. In addition, associated major and minor anomalies can have an important impact on whether cases are included or not included in a study and, until molecular definitions begin to separate what should or should not be included in a particular definition, the discussion and criteria need to be established on the basis of clinical and embryologic grounds. Some definitions of non-syndromic clefting disorders would exclude any child with any other major organ system malformation, as well as a number of minor malformations, while other systems might allow the inclusion of a single major, or one or two minor, malformations. Recent developments in ultrasound also afford the opportunity to look for sub-clinical manifestations of clefts, such as deficiencies of the orbicularis oris muscle; these can also be very valuable tools for generating such sub-phenotypes.

The description of what constitutes non-syndromic forms of clefting ... has yet to be fully resolved

In conclusion, and discussed further in Section 6 below, it is clear that these issues need to be formally addressed in any study that is carried out, and that investigators engaged in collaborative studies need to have consensus views for case inclusion and exclusion. Until the molecular phenotypes begin to help sort this out, both narrowly as well as broadly defined phenotypes may be used in genetic mapping studies; the availability of powerful computer analytic programmes also affords the opportunity to carry out multiple sets of analyses on subsets of clinically-defined cases, all drawn from a common larger data set. Table 8, below, shows some disorders where affected individuals might present as a "non-syndromic" cleft.

Table 8: Single-gene disorders that can mimic non-syndromic clefting

Phenotype	Single-gene disorder
Cleft lip and/or palate	CPX
Cleft lip and/or palate	EEC
Cleft lip and/or palate	CLPED1
Cleft lip and/or palate	VDWS

Source: Murray JC, 2002

5.3 Mouse models

The utility of the mouse for comparative studies of human genetic disorders has been widely acknowledged since the early 1900s. This work has become even more valuable as the ability to generate gene-specific knockouts or over-expression transgenics has become available. Coupled to the utility of the mouse as an experimental organism in which embryo manipulation can be carried out, is the very powerful genetics available through this system in which many generations of controlled breeding can be performed in a relatively short period of time. Finally, since the mouse is a mammal, many of its embryologic and developmental processes are closely related to those of the human. In the area of craniofacial development, studies of the mouse have been especially productive. A large number of knockout and transgenic animals that have been generated demonstrate disruptions of craniofacial structures and have provided opportunities to investigate genes identified in development. Evaluation of genes whose expression pattern also supports a role for development of craniofacial structures has also been critical. Particularly relevant models in the mouse come from knockouts of MSX1 (Satokata and Maas, 1994), TGFβ 3 (Proetzel et al., 1995) and SKI (Colmenares et al., 2002). Spontaneously arising mutations, particularly ones in which the defects are focused on a specific craniofacial structure, such as the cleft models CLF1 and 2 studied by Diana Juriloff (2001), have also been particularly relevant. And finally, the work of investigators, such as Robert Erickson and Scott Diehl (1997), in carrying out genome-wide strategies to look at gene/environment interactions and the role of teratogens in mouse models of clefting has also been very fruitful in providing localizations to regions that have high homology to human chromosomes as a way to better understand these forms of interactions. The availability of large amounts of mouse DNA sequence and very detailed mouse genetic maps and reagents for carrying out mapping also make the mouse an especially productive engine for the study of craniofacial anomalies. During the course of the meeting, details as well as new data were presented by Drs Diehl, Erickson and Juriloff and

provided opportunities for investigators working in human genetic systems to interact directly.

5.4 Genotyping

Advances as well as current strategies revolving around the issue of genotyping were discussed, particularly as they relate to humans. Genotyping includes the genetic analysis of variation and, for purposes of studies of cleft lip and palate, can be applied to genome-wide searches for gene or locus identification or to association studies using candidate-gene analysis. In addition, discussion about the use of chromosomal anomalies in gene finding was also provided. Besides the methodologies involved in the genotyping *per se*, discussions over strategies and particular analytic approaches were also carried out.

5.4.1 Strategic approach

BOX 1

Strategic approaches

The cleft lip and palate genetics literature is a fusion of studies that have made use of candidate-gene and association analysis with a more limited number of studies that have used a linkage or genome-wide approach. A recent review summarized the "state-of-the-field" in 2002 (Murray, 2002) with loci on chromosomes 1, 2, 4, 6 and 14 holding particular interest. The difficulties in studying a complex disease, such as cleft lip and palate, include identification of a sufficient number of families in order to have the ability to effectively carry out a genome-wide approach. Thus, many early studies as well as current studies have made use of candidate-gene approaches to look at cases and have compared allelic frequencies with a control population. Very recently the ability to carry out direct candidate-gene sequencing has also been incorporated into some studies. While no single approach is likely to provide all the answers, there have been some preliminary successes with each of the above-mentioned approaches. In addition to the strategic approach selected, individual methodologies are also rapidly changing – as is common in molecular biology – and ongoing evaluation of the specific methodologic approaches will also be key for projects selected within individual laboratories. Finally, the ability to coordinate either analytic techniques or specific methodology and marker selection were important issues also discussed.

5.4.2 Analysis

Analytic approaches can use, for genome-wide searches, either parametric or non-parametric analyses. High-density genetic maps (Broman et al., 1998) and public resources for genotyping such as the NIH-sponsored Centre for Inherited Disease Research (CIDR) provide opportunities for even modestly-funded investigators to undertake such searches. Parametric analysis is the standard linkage approach in which the mechanism of inheritance pattern needs to be specified and can greatly benefit when a single large family is available. A recent report by van den Boogaard (2000) illustrates the utility of this approach when a single large family segregating for cleft lip and palate was identified, shown to be linked to the MSX1 locus on chromosome 4, and a point mutation resulting in a stop codon within this gene eventually identified. This family is especially remarkable in that many of the individuals have a phenotype that, if viewed in isolation, would be readily characterized as non-syndromic cleft lip and palate and raises the possibility of this disorder, at least in some cases, being caused by mutations in MSX1. The difficulty of this approach is that large families, such as the one described by van den Boogaard, are rare and may not provide insights into the most important or frequent genes involved in non-syndromic forms of clefting. Pools of such families can also be used in standard linkage analysis, and this approach has been used for many other complex disorders.

Within ... the parametric and non-parametric categories there are many competing analytic strategies

A compliment to the parametric approach is the non-parametric approach or the affected-pedigree member technique. This approach is best exemplified by sib-pair analysis in which pairs of affected siblings are identified and evidence for statistical aberrations in the proportion of alleles shared either by identity or descent established through genotyping. This approach, though powerful in that genetic mechanisms do not have to be specified, is unable to provide the more defined locus identification that will come about through linkage approaches. Most investigators would now choose to assemble a collection of families in which either analytic strategy, in general, could be applied and then carry out complimentary analysis using a multiplicity of approaches. Even within each of the broad categories – parametric and non-parametric – there are many competing analytic strategies that are discussed in more detail in other publications. Furthermore, the addition of analysis of variance approaches in which the severity of the phenotype can be taken into account, as well as the addition of environmental variables as an analytic variable, are also important considerations in current study designs (Almasy and Blangero, 1998). Although only one large genome-wide search has been carried out (Prescott et al., 1999), there are now under way genome-wide approaches from other laboratories; it is likely that over the next few years additional evidence from these searches will be provided.

Several candidate loci searches using 10 to 40 families have already been reported (Carinci et al., 2000; Marazita et al., 2002). Table 9 below summarizes some linkage work done in humans.

Table 9: Human loci/genes for suggested linkage

Position	Disorder*	Method**	Cloned	[Candidate]
1p36	NS	L/CH	–	[SKI]
1q32	VDWS	L/CH	IRF6	–
2p13	NS	L/LD	TGFα	–
3q27	EEC3	L/KO	P63	–
4p16	NS	LD/CH/KO	MSX1	–
4q31	NS	L/LD	–	–
6p23	NS	L/CH	–	[AP2, EDN1]
11q23	ED4	L	–	–
14q24	NS	LD/KO	TGFβ 3	–
9q13	NS	L/LD	–	–
Xq21	CPX	L/CH	TBX22	–

Source: Murray JC, 2002

* Disorders:
NS: non syndromic
VDWS: Van der Woude syndrome
EEC3: ectodermal dysplasia/ectrodactyly and clefting syndrome 3
ED4: ectodermal dysplasia and clefting syndrome 4
CPX: X-linked cleft palate and ankyloglossia

** Methods:
L: linkage
CH: chromosomal rearrangement
LD: linkage disequilibrium
KO: mouse knockout

The complements to family-based approaches are those that use case and control populations. These studies are best carried out when a candidate gene or locus is available as they depend on the phenomenon of linkage disequilibrium, active over only very short physical distances of DNA. This limits the study to a handful of 10 to perhaps 100 loci, given current fiscal realities and available markers. The selection of candidate genes can often take place using the descriptions provided through developmental biology or mouse models, and frequently utilizes genes shown to be expressed in the developing palate or genes whose disruption in a knockout mouse, for example, would result in a cleft lip or palate phenotype. Judicious selection of candidate genes can be an effective tool in identifying a genetic component of a common disorder. These studies in cleft lip and palate were initiated in 1989 with the study of Ardinger et al. in which evidence for the role of TGFα was provided and a case-control approach was

followed, using non-syndromic cleft lip and palate as the cases and convenience controls, selected from the same geographic area as the cases were collected. Since this publication, the literature has expanded greatly with a number of additional studies, including those using more powerful analytic techniques, that have provided both positive and negative results. A summary of these studies, given in Table 10 below, would seem to support some evidence that both TGFβ 3 and MSX1 are genes involved in clefting.

Table 10: Candidate-gene studies for CL/P

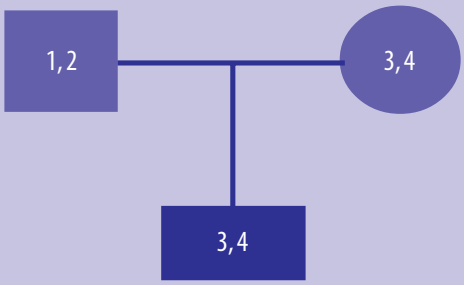
Gene	Analysis	Result	Author
TGFα	meta	OR=1:43 (1.12 to 1.80)	Mitchell
MSX1	case-control	p<0.005	Lidral
MSX1	AFBAC*	p<0.04	Lidral
MSX1	TDT	p<0.001	Vieira
TGFβ 3	TDT	p<0.008	Lidral
TGFβ 3	TDT	p<0.01	Maestri
TGFβ 3	TDT	p<0.02	Vieira

* Affected family member-based controls.

Source: Murray JC, 2002

The addition of newer analytic strategies, such as the transmission disequilibrium test (TDT) (Spielman and Ewens, 1996) and likelihood ratio test (LRT) (Umbach and Weinberg, 2000) tests in which transmission distortion or family-based allelic controls to prevent the confounding of ethnic matching, provides for even more powerful platforms for the collection of information. In addition, it is now possible and feasible to collect hundreds of families with a focus on nuclear triads, consisting of an affected child with the mother and father (as shown below in Figure 1 below), in which substantial power for detecting even small gene effects is available.

Figure 1: Nuclear families as internal controls



As the selection of candidate-gene panels also becomes more robust, these approaches are likely to be successful.

5.4.3 Sample collection

Sample collection issues are of paramount importance and were discussed widely. While it is easy to collect samples in the form of buccal or cheek swabs, for example, the DNA available from these is limited and, at the present time, is unlikely to comprise enough for a genome-wide search. Whole blood samples are more robust, both in terms of the quality and quantity of DNA available, and are usually sufficient to apply to genome-wide searches in which approximately 20 micrograms of high quality DNA would be required. Whole blood, however, can present challenges in collection and, in the case of small infants, may be limited by available quantities. Additional advantages of whole blood include the possibility of saving plasma or serum for analysis of other analytes, such as micronutrients or storing cells for subsequent RNA or protein studies. Other tissues, including cord blood, placenta and materials obtained from the site of surgery, also provide opportunities for other forms of analysis. Materials obtained at the site of surgery, for example, might be useful for looking at abnormalities of gene expression found in affected tissues. While there is no single sample collection strategy that can solve all the financial and technical problems, the issues raised by these were important considerations for the group as a whole and, from ongoing studies, it is clear that a variety of study designs have been selected as most appropriate for particular projects. For example, the large collaborative study under way, sponsored by the US Centres for Disease Control, has chosen buccal swabs as these can be obtained via the mail from individuals who self-collect on themselves and their children. The advantage is that this is a very cost-effective approach; it allows for the collection of thousands of samples yearly on a limited budget and also provides limited amounts of DNA for analysis. Other studies have collected blood-spot samples and these may prove to be especially effective when studying newborn populations. These samples are useful in that they can be stored indefinitely and inexpensively but are compromised by the limited quality and amount of their DNA; there may be challenges to comprehensive analysis of DNA from such samples where only certain genotyping approaches may work.

5.4.4 Collaborative strategies

See Box J, facing page.

BOX J

Collaborative strategies

A variety of efforts, already under way, foster collaborative interactions in the study of cleft lip and palate. A few of these are described below.

1. Estudio Colaborativo Latino Americano Malformaciones Congenita (ECLAMC)

ECLAMC is a collaboration, established in the mid-1960s in South America, in which up to 100 participating hospitals have one or more volunteer paediatricians who collect demographic and clinical data on a wide spectrum of structural birth defects, including cleft lip and palate. The data is entered in a common format and returned to a central repository for storage and evaluation. While this collaboration makes use of volunteer physicians, it has proven to be highly effective and currently collects data on approximately 200 000 cases per year. Numerous studies have been published by this group, including some relevant to cleft lip and palate suggesting, for example, that altitude or ethnicity may be important roles in determining risks for clefts. Recently the group has also incorporated blood sampling from children and parents into their strategy, and it is likely that this will provide extremely powerful data for analyses, given the large number of samples available.

2. US Centers for Disease Control and Prevention (CDC)

For the last five years, the US Centers for Disease Control and Prevention has sponsored an eight-location collaboration to collect data from 2400 cases and 800 controls per year, with a collection of 30 structural birth defects, including cleft lip and palate. In addition, biological sample collection in the form of cheek swabs, collected from infants and their parents, has also been incorporated to complement an extensive interview of the mother in which data regarding pregnancy risks, such as drug exposures, outcomes, nutritional factors and family history, are all incorporated. Data, as well as biological samples, are stored in a central repository and made available to collaborating investigators for addressing specific hypotheses. Because there are such detailed characterizations of environmental exposures along with the collection of DNA samples, this project has enormous power to study gene/environment interactions across a broad geographic range in the United States.

3. European Collaboration on Craniofacial Anomalies (EUROCRAN)

A multi-centre collaboration funded by the European Union (Contract Number: QLG1-CT-2000-01019) was established, combining existing networks that have already been established by EUROCLEFT and the European Science Foundation (ESF) (http://www.esf.org). A pan-European, multi-centre, multidisciplinary effort has evolved. The innovation arises from the involvement of international experts at the cutting edge of research in their respective fields, and the application of advances in basic sciences and molecular biology to clinical research is seen as the way forward. A number of ground-breaking work packages have been undertaken, collectively aimed at improving knowledge on the etiology and pathogenesis of craniofacial abnormalities, introducing precise diagnostics/risk assessment, developing therapeutics and producing the best (evidence-based) treatment protocols. These research efforts are being extended to Eastern Europe, and the ultimate objective is to pursue their implementation further afield *(see Annex 1)*.

4. European Registry for Congenital Anomalies and Twins (EUROCAT)

EUROCAT is a European network of registries for the epidemiologic surveillance of congenital anomalies. EUROCAT began in 1979 and currently surveys more than 900 000 births per year. Through its work on harmonization of methodology, particularly for ascertainment, EUROCAT has become an established reference centre for population-based information on congenital anomaly prevalence and time trends. The EUROCAT collaborative framework seeks to exploit the power of transnational collaboration in data collection and exchange of expertise to address issues of concern on birth-defects prevention and service delivery (http://www.lshtm.ac.uk/php/eeu/eurocat).

5.5 Recent developments

While the field of craniofacial anomalies and genetic studies is rapidly moving, a few comments about recent developments are useful. Genes continue to be cloned for a variety of syndromic forms of cleft lip and palate and, very recently, the first craniofacial anomaly identified through linkage – X-linked cleft palate/ankyloglossia syndrome – has had its gene (TBX22) identified (Braybrook et al., 2001). In this case, a transcription factor, TBX22, has been shown to be at fault, and this further opens the door for additional investigations of other transcription factors or their pathway members in non-syndromic forms of clefting. In a complimentary report, the Spritz group (Sozen et al., 2001) has provided evidence that heterozygotes for the PVRL1 gene, which had previously been shown to have etiologic mutations in the Margarita Island ectodermal dysplasia clefting syndrome (Suzuki et al., 2000), had heterozygotes that have an increased frequency of non-syndromic clefting in populations studied in Venezuela. This raises the possibility that heterozygotes for syndromic forms of clefting might occasionally be at increased risk for non-syndromic forms and that, potentially, gene/environment interactions might further complicate this story. This is an important and exciting finding that opens the door to many additional forms of investigation. Candidate-gene studies have continued to be expanded and Terri Beaty's group (2002) has also recently reported additional evidence for the role of the MSX1 homeobox gene in cleft lip and palate. The gene for the van der Woude and Popliteal pterygium syndromes, interferon regulatory factor 6 (IRF6), has also been reported (Kondo et al., submitted). Finally, new efforts at genome-wide approaches are under way and are likely to contribute new information in the near future.